

# Student Homework Sheet — Stage 11: Evaluation & Risk Communication

---

## Chain

In the lecture, we learned how to compare **parametric vs bootstrap CIs**, run **scenario sensitivity**, and present **assumption-aware** results with subgroup checks.

Now, you will adapt those methods to your dataset to quantify uncertainty, compare at least two assumptions, and communicate risks clearly.

## Task Overview

You will:

1. Run a **bootstrap** to estimate uncertainty for a prediction or metric.
2. Compare **two “what-if” scenarios** (e.g., imputation choice, distributional assumption, or model configuration).
3. Produce **visuals** (CI bands, error bars, residual plots, side-by-side scenarios, etc.).
4. Write a **stakeholder-facing summary** stating assumptions, risks, and how results change.

## Step-by-step

### 1. Load data & model

- Use your dataset or the provided synthetic fallback.
- Starter files:
  - `/notebooks/stage11_evaluation-risk-communication_homework-starter.ipynb`
  - `/src/evaluation.py` (helper functions)
  - `/data/data_stage11_eval_risk.csv` (auto-created if missing)

### 2. Baseline fit

- Fit your model and compute at least one metric (RMSE, MAE, AUC, etc.).

### 3. Bootstrap

- Resample your data  $\geq 500$  times to estimate a **confidence interval** for the chosen metric or prediction.
- Visualize uncertainty with **CI bands, error bars, or other relevant plots**.

### 4. Scenario comparison ( $\geq 2$ )

- Examples: mean vs median imputation, Gaussian vs t-distribution noise, drop vs fill, linear vs polynomial fit.
- Show how results shift and discuss the implications.

### 5. Subgroup diagnostics

- Pick a categorical split (e.g., `segment`) and compare residuals/metrics.

- Identify hidden failures or confirm stability across subgroups.

## 6. Visualize

- Side-by-side panels for scenarios and subgroups with captions: *Assumptions* and *Takeaway*.
- Ensure axes are consistent for comparability.

## 7. Write-up ( $\leq 1$ page)

- State assumptions, describe risks, summarize sensitivity results.
- Clearly communicate **when the model is reliable vs risky**.
- Use plain language for stakeholders, e.g., *"Prediction holds if weekly volatility stays within X; model sensitive to missing-rate > 10%; Segment C underperforms."*

## Deliverables

- Notebook in `/notebooks/` named `stage11_eval_risk_homework.ipynb` containing:
  - Bootstrap code & CI figure
  - Scenario comparison visuals
  - Subgroup diagnostics
  - Markdown summary (assumptions, risks, sensitivity)
- Save any helper functions in `/src/evaluation.py` if you modify or add functions.

**Submission:** Save your completed notebook in `/notebooks/` and push to your repository.

## Rubric (10 points)

- **Reproducibility (2 pts):** seed set, clear steps, functions reused.
- **Bootstrap CI (2 pts):** correct resampling & CI computation; interpretation is sound.
- **Scenarios (3 pts):**  $\geq 2$  clearly different assumptions; side-by-side visuals; consistent axes; labels.
- **Subgroup diagnostic (1 pt):** evidence of heterogeneity or confirmation of stability.
- **Stakeholder write-up (2 pts):** plain-language assumptions, "holds if...", "sensitive to...", risks and next steps.

## Example Expectation

- A  $2 \times 2$  grid: Baseline vs Scenario A vs Scenario B with CIs, plus a residuals-by-segment chart.
- A concise paragraph summarizing key insights for stakeholders.

### Notes:

- Connect your homework to the lecture content: parametric vs bootstrap CIs, scenario comparisons, and subgroup evaluation.
- Use all visuals and tables to communicate uncertainty and risk clearly.
- Document all assumptions and any deviations from starter notebook code.